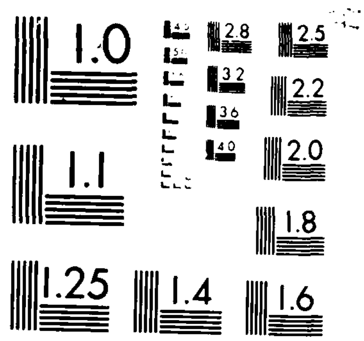
AD-A177 197 SOME RECENT RESULTS IN SIGNAL DETECTION(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS C R RAO SEP 86 TR-86-23 AFOSR-TR-87-0260 F49620-85-C-0008 1/1

UNCLASSIFIED F/G 9/4 NL




MICROCOPY RESOLUTION TEST CHART

# SOME RECENT RESULTS IN SIGNAL DETECTION

BY

C. Radhakrishna Rao

Center for Multivariate Analysis
University of Pittsburgh
Pittsburgh, PA 15260



**Center for Multivariate Analysis**

**University of Pittsburgh**

(2)

# SOME RECENT RESULTS IN SIGNAL DETECTION

BY

C. Radhakrishna Rao

Center for Multivariate Analysis
University of Pittsburgh
Pittsburgh, PA 15260

September 1986

Technical Report 86-23

Center for Multivariate Analysis
515 Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260



This work is supported by Contract N00014-85-K-0292 of the Office of Naval Research and Contract F49620-85-C-008 of the Air Force Office of Scientific Research. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

# SOME RECENT RESULTS IN SIGNAL DETECTION*

by

C. Radhakrishna Rao**

Center for Multivariate Analysis
University of Pittsburgh
Pittsburgh, PA 15260

ABSTRACT

Some recent results on the detection and estimation of signals in the presence of noise are discussed. An exact confidence lower bound is obtained for the discriminatory power of an estimated linear discriminant function for signal detection. Information theoretic criteria are suggested for the estimation of number of signals. A new method is proposed for determining the number of signals and estimating them in exponential signal models.

Key Words and Phrases: Discriminant function, exponential signal models, information criteria in model selection, Prony's method, signal processing.

* Paper presented at the Fourth Purdue Symposium, at W. Lafayette, Indiana, June 15-20, 1986.

** Research sponsored by the Air Force Office of Scientific Research (AFSC) under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 87-0260
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): Some recent results in signal detection
5. TYPE OF REPORT & PERIOD COVERED: Technical - September 1986
6. PERFORMING ORG. REPORT NUMBER: 86-23

7. AUTHOR(s): C. Radhakrishna Rao
8. CONTRACT OR GRANT NUMBER(s): F49620-85-C-0008

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Center for Multivariate Analysis
University of Pittsburgh
512 Thackeray Hall, Pittsburgh, PA 15260

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F, 2304

11. CONTROLLING OFFICE NAME AND ADDRESS:
Air Force Office of Scientific Research
Dept. of the Air Force
Bolling Air Force Base, DC 20332

12. REPORT DATE: September 1986
13. NUMBER OF PAGES: 22

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same as 11
15. SECURITY CLASS. (of this report): unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):
Discriminant function, exponential signal models, information criteria in model selection, Prony's method, signal processing.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):
Some recent results on the detection and estimation of signals in the presence of noise are discussed. An exact confidence lower bound is obtained for the discriminatory power of an estimated linear discriminant function for signal detection. Information theoretic criteria are suggested for the estimation of signals. A new method is proposed for determining the number of signals and estimating them in exponential signal models.

DD FORM 1 JAN 73 1473

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# 1. INTRODUCTION

Some of the problems of signal detection considered by electrical engineers can be described in terms of a linear model

$$Y = \xi + \eta \tag{1.1}$$

where $\xi$ is an n-vector signal, $\eta$ is an n-vector noise variable perturbing the signal during transmission and Y is the received message. The general problem associated with the model (1.1) is that of detecting the presence of a signal in a received message and estimating it when present. Different types of problems arise depending on the nature of $\xi$ and $\eta$.

We consider two types of problems, one where $\xi$ is considered as a specified vector or a specified function of unknown parameters, and another where $\xi$ is considered as a stochastic vector distributed independently of the noise vector $\eta$.

The following notations are used. A' denotes the transpose of a matrix A when its elements are real and A* the conjugate transpose of A when its elements are complex.

(i) $X \sim N_p(\mu,\Sigma)$, i.e., a real p-vector X has a p-variate real normal distribution with the probability density function (p.d.f.)

$$(2\pi)^{-p/2}|\Sigma|^{-\frac{1}{2}}\exp\left[-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right]. \tag{1.3}$$

(ii) $X \sim \tilde{N}_p(\mu,\Sigma)$, i.e., a complex vector X has a p-variate complex normal distribution with the p.d.f.

$$(\pi)^{-p}|\Sigma|^{-1}\exp\left[-(x-\mu)^*\Sigma^{-1}(x-\mu)\right]. \tag{1.4}$$

(iii) $S \sim W_p(f,\Sigma)$, i.e., a real p x p positive definite matrix S has the Wishart distribution on f degrees of freedom with the p.d.f.

$$2^{-pf/2} [\Gamma_p(f/2)]^{-1}|\Sigma|^{-f/2}|S|^{(f-p-1)/2}\exp(-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}S) \tag{1.5}$$

where

$$\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=1}^{p} (a- \frac{i-1}{2} ).$$

(iv) $S \sim \tilde{W}_p(f,\Sigma)$, i.e., a complex p x p positive definite matrix S has the complex Wishart distribution with the p.d.f.

$$[\tilde{\Gamma}_p(f)]^{-1}|\Sigma|^{-f}|V|^{f-p}\exp(-\mathrm{tr}\Sigma^{-1}S) \tag{1.6}$$

where

$$\tilde{\Gamma}_p(a) = \pi^{p(p-1)/2} \prod_{i=1}^{p} (a-i+1).$$

## 2. PROBLEMS INVOLVING A FIXED SIGNAL

### 2.1 Test for a specified signal

Consider the model

$$X = \xi + \eta \tag{2.1.1}$$

where $\xi$ is a fixed vector and $\eta \sim N_p(0,a^{-1}\Sigma)$ with a known scalar a and an unknown covariance matrix $\Sigma$. Let S be a p×p positive definite matrix variable such that $S \sim W_p(f,\Sigma)$. We have independent observations on X and S, on the basis of which we wish to test the hypothesis

$$H_0 : \xi = 0 \quad \text{versus} \quad H_1 : \xi = \delta \text{ (specified)}. \tag{2.1.2}$$

Rejection of $H_0$ at a chosen level of significance would indicate that the received message X contains the signal $\delta$ and is not pure noise.

It may be noted that when $p = 1$, the appropriate test of $H_0$ is the one-sided t test

$$t = \frac{a^{\frac{1}{2}} X}{(s/f)^{\frac{1}{2}}} > c \text{ if } \delta > 0 \text{ (or } < c \text{ if } \delta < 0) \tag{2.1.3}$$

on f degrees of freedom. When $p > 1$, one generally uses Hotelling's

$$T^2 = \frac{a(f-p+1)}{p} X'S^{-1}X \tag{2.1.4}$$

which is distributed as F on p and $f-p+1$ degrees of freedom. The test (2.1.4), however, does not involve the specified $\delta$. A more powerful test than (2.1.4) is recently suggested by Khatri and Rao (1985b) based on the following considerations.

Let C be a $p\times(p-1)$ matrix of rank $p-1$ such that $\delta'C = 0$ and consider the transformation

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \delta' \\ C' \end{pmatrix} X \ , \ V = \begin{pmatrix} \delta' \\ C' \end{pmatrix} S \ (\delta : C). \tag{2.1.5}$$

Then

$$Y \sim \begin{cases} N_p\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, a^{-1}\Sigma_*\right) \text{ under } H_0 \\ N_p\left(\begin{pmatrix} \delta'\delta \\ 0 \end{pmatrix}, a^{-1}\Sigma_*\right) \text{ under } H_1 \end{cases}$$

$$V \sim W_p(f, \Sigma_*) \ , \ \Sigma_* = \begin{pmatrix} \delta' \\ C' \end{pmatrix} \Sigma(\delta : C) \tag{2.1.6}$$

and the problem (2.1.2) reduces to testing

$$H_0 : E(Y_1) = 0 \quad \text{versus} \quad H_1 : E(Y_1) = \delta'\delta > 0$$
$$\text{given } E(Y_2) = 0. \tag{2.1.7}$$

Such a problem involving a conditional test was considered in Rao (1946). The appropriate test is $t > t_\alpha$ where

$$t = \frac{a^{\frac{1}{2}}(f-p+1)^{\frac{1}{2}}\delta' S^{-1} X}{[(1+aX'S^{-1}X)\delta' S^{-1}\delta - a(\delta' S^{-1} X)^2]^{\frac{1}{2}}} \tag{2.1.8}$$

has t distribution on (f-p+1) degrees of freedom and $t_\alpha$ is the upper tail $\alpha$% point of the t distribution.

In testing (2.1.7), we used the condition $E(Y_2) = 0$. But in practice it may be necessary to check whether this holds. For this, we use Hotelling's

$$T^2 = \frac{a(f-p+2)}{p-1}\left(X'S^{-1}X - \frac{(\delta' S^{-1} X)^2}{\delta' S^{-1}\delta}\right) \tag{2.1.9}$$

which has F distribution on (p-1) and (f-p+2) degrees of freedom.

The tests (2.1.8) and (2.1.9) can be extended to the case where $\eta$ and S have complex normal and Wishart distributions respectively. For example, the test for $H_0:\mu=0$ versus $H_1:\mu=\delta$ is $t \geq t_\alpha$ where

$$t = \text{real part of } \frac{[2(f-p+1)]^{\frac{1}{2}}\delta^* S^{-1} X}{[(a^{-1}+X^*S^{-1}X)\delta^* S^{-1}\delta - \delta^* S^{-1} X X^* S^{-1}\delta]^{\frac{1}{2}}} \tag{1.2.10}$$

has t distribution on 2(f-p+1) degrees of freedom. Further the hypothesis that $E(Y_2) = 0$ is tested by Hotelling's

$$T^2 = \frac{a(f-p+2)}{p-1} \left(X^*S^{-1}X - \frac{\delta^*S^{-1}XX^*S^{-1}\delta}{\delta^*S^{-1}\delta}\right) \qquad (2.1.11)$$

which has F distribution on 2(p-1) and 2(f-p+2) degrees of freedom.

## 2.2 Discrimination between noise and a specified signal

In Section 2.1, we considered the problem of testing for pure noise against a specified signal on the basis of an observed message X and an independent estimate $f^{-1}S$ of $\Sigma$. The interpretation of such a test at a chosen level of significance is not simple, specially when the same estimate $f^{-1}S$ of $\Sigma$ is used repeatedly to test for signals in a number of incoming future messages. In order to provide a satisfactory solution, we consider the problem of signal detection as one of discrimination between alternative populations $N_p(0,\Sigma)$, (or $\tilde{N}_p(0,\Sigma)$), and $N_p(\delta,\Sigma)$, (or $\tilde{N}_p(\delta,\Sigma)$), when $\Sigma$ is unknown but an estimate $f^{-1}S$ of $\Sigma$ is available. In such a case, the estimated linear discriminant function (LDF) is proportional to

$$y = \delta' S^{-1}X, \text{ (or } \delta^* S^{-1}X). \qquad (2.2.1)$$

[In the sequel we follow the practice of giving the results for the real case first and then for the complex case within brackets as in (2.2.1)]. The discriminatory power of (2.2.1) when applied to future observations is a monotone function of the discrimination index (DI)

$$\rho(S,\Sigma) = \frac{[E(y|S, \xi=\delta) - E(y|S,\xi=0)]^2}{V(y|S)}$$

$$= \frac{\delta' S^{-1}\delta}{\delta' S^{-1}\Sigma S^{-1}\delta}, \text{ (or } \tilde{\rho}(S,\Sigma) = \frac{\delta^* S^{-1}\delta}{\delta^* S^{-1}\Sigma S^{-1}\delta}). \qquad (2.2.2)$$

If $\Sigma$ is known, the true LDF is $\delta'\Sigma^{-1}X$, (or $\delta^*\Sigma^{-1}X$), and the optimal DI is

$$\rho(\Sigma,\Sigma) = \delta'\Sigma^{-1}\delta, (\text{or } \tilde{\rho}(\Sigma,\Sigma)=\delta^*\Sigma^{-1}\delta) \qquad (2.2.3)$$

which is greater than or equal to (2.2.2) by the Cauchy-Schwartz inequality, so that there is, in general, loss of information in using an estimated $\Sigma$. The expression (2.2.2) involves the unknown covariance matrix $\Sigma$, and therefore, the realized discriminatory power by using a particular estimate of $\Sigma$ in classifying future observations remains unknown. We raise the question as to whether an estimate of (2.2.2) can be obtained in terms of known (observed) values to provide some idea of the performance of the estimated LDF. One such is the plug-in estimate $\rho(S,f^{-1}S)$, for $\tilde{\rho}(S,f^{-1}S))$, but it is known to be highly biased estimator of (2.2.2). In a recent paper, Khatri and Rao (1985a) provided a satisfactory solution, which is as follows.

It is shown that in the real case

$$B = \frac{(\delta'S^{-1}\delta)^2}{(\delta'\Sigma^{-1}\delta)(\delta'S^{-1}\Sigma S^{-1}d)} \quad \text{and } G = \frac{\delta'\Sigma^{-1}\delta}{\delta'S^{-1}\delta} \qquad (2.2.4)$$

are independently distributed with p.d.f. (probability density function) of B as

$$\frac{\Gamma(\frac{f+1}{2})}{\Gamma(\frac{f-p+2}{2})\ \Gamma(\frac{p-1}{2})} b^{(f-p)/2}(1-b)^{(p-3)/2} \qquad (2.2.5)$$

and that of G as

$$\frac{1}{2^{(f-p+1)/2}\ \Gamma(\frac{f-p+1}{2})} e^{-g/2}\ g^{(f-p+1)/2}. \qquad (2.2.6)$$

In the complex case, defining $\tilde{B}$ and $\tilde{G}$ with $\delta'$ replaced by $\delta^*$ in (2.2.4), it is shown that $\tilde{B}$ and $\tilde{G}$ are independently distributed with the p.d.f. of $\tilde{B}$ as

$$\frac{\Gamma(f+1)}{\Gamma(f-p+2)\,\Gamma(p-1)}\,\tilde{b}^{(f-p+1)}(1-\tilde{b})^{p-2} \tag{2.2.7}$$

and that of $\tilde{G}$ as

$$\frac{1}{\Gamma(f-p+1)}\,e^{-\tilde{g}}\,\tilde{g}^{(f-p)}. \tag{2.2.8}$$

The distribution (2.2.7) was earlier obtained by Reed, Mallet and Brennan (1974). The distributions (2.2.3) - (2.2.8) are independent of the unknown parameters which enables us to draw inferences on $\rho(S,\Sigma)$, (or $\tilde{\rho}(S,\Sigma)$), through the pivotal quantities $\tilde{B}$ and $\tilde{G}$, (or B and G) as discussed below.

1. Using the expressions for the moments of the beta distribution (Rao (1973), p. 168)

$$E(B \text{ or } \tilde{B}) = \frac{f-p+2}{f+1}$$

we have

$$E[\rho(S,\Sigma)),(\text{or } \tilde{\rho}(S,\Sigma))] = \frac{f-p+2}{f+1}\,[\delta'\Sigma^{-1}\delta,(\text{or } \delta^*\Sigma^{-1}\delta)]. \tag{2.2.9}$$

If the average efficiency is to be maintained at about half the optimal efficiency, then from (2.2.9) we have

$$\frac{f-p+2}{f+1} \cong \frac{1}{2} \text{ or } f \cong 2p \tag{2.2.10}$$

for both real and complex cases, i.e., the degrees of freedom on which $\Sigma$ is estimated should be at least twice the number of components of the signal.

This result for the complex case is mentioned in Reed, Mallet and Brennan (1974). Similarly we can equate the ratio in (2.2.9) to any desired ratio other than (1/2) and find the degrees of freedom f needed for the estimate of $\Sigma$.

2. Perhaps, a more satisfactory way of using the distributions (2.2.4) and (2.2.7) is as follows. Let $b_\alpha$, (or $\tilde{b}_\alpha$), be the lower $\alpha$% point of the distribution (2.2.4), (or (2.2.7)). Then we can make the confidence statement that

$$\rho(S,\Sigma) \geq b_\alpha \delta'\Sigma^{-1}\delta, \text{ (or } \tilde{\rho}(S,\Sigma) \geq \tilde{b}_\alpha \delta^*\Sigma^{-1}\delta) \qquad (2.2.11)$$

with a confidence coefficient of $(1-\alpha)$%.

3. The results (2.2.10) and (2.2.11) still involve the unknown quantity $\Sigma$. We raise the question, whether the actual magnitude of $\rho(S,\Sigma)$ for given S can be assessed through known values. Using the joint distribution of B and G as in (2.2.5) and (2.2.6) it is shown in Khatri and Rao (1985a) that $E[\rho(S,\Sigma)-gD^2]^2$ attains its minimum at $g = (f-p+2)(f-p-1)/f(f+1)$ so that

$$\hat{\rho}(S,\Sigma) = \frac{(f-p+2)(f-p-3)}{f(f+1)} D^2 \qquad (2.2.12)$$

which does not involve unknown parameters, in a close approximation to $\rho(S,\Sigma)$. Similarly

$$\hat{\tilde{\rho}}(S,\Sigma) = \frac{(f-p+2)(f-p-1)}{f(f+1)} D^2 \qquad (2.2.13)$$

is a close approximation to $\tilde{\rho}(S,\Sigma)$.

4. We can also obtain the exact lower confidence bound to $\rho(S,\Sigma)$, which provides a satisfactory answer to the problem raised. We define the random variable

$$Z = \frac{1}{2} BG = \frac{f}{2} \frac{\rho(S,\Sigma)}{D^2} \text{ (or } \tilde{Z} = \frac{f\tilde{\rho}(S,\Sigma)}{D^2}\text{ )} \qquad (2.2.14)$$

which has the confluent hypergeometric distribution with the p.d.f.

$$\frac{e^{-z} z^{m-1}}{\Gamma(m)} \frac{\Gamma(a+b)}{\Gamma(a)} \Psi(b,m-a+1;z) \qquad (2.2.15)$$

where $m = (f-p+1)/2$, $a = (f-p+2)/2$, $b = (p-1)/2$, (or $m = f-p+1$, $a = f-p+2$, $b = p-1$), and

$$\Psi(b,c;z) = \frac{1}{\Gamma(b)} \int_0^{\infty} t^{b-1} (1+t)^{c-b-1} \exp(-zt)dt. \qquad (2.2.16)$$

The percentage points of this distribution at various levels are tabulated in Khatri, Rao and Sun (1986). If $z_\alpha$ (or $\tilde{z}_\alpha$) is the lower $\alpha$% point of this distribution, then

$$\rho(S,\Sigma) \geq \frac{2z_\alpha}{f} D^2 \text{ (or } \tilde{\rho}(S,\Sigma) \geq \frac{\tilde{z}_\alpha}{f} D^2) \qquad (2.2.17)$$

provides the lower confidence bound to $\rho(S,\Sigma)$, (or $\tilde{\rho}(S,\Sigma)$), with a confidence coefficient of $(1-\alpha)$%.

Several approximations to the distribution (2.2.15) are also obtained in Khatri, Rao and Sun (1986) when f is large compared to p, from which fairly accurate percentage points can be easily obtained.

## 3. PROBLEMS INVOLVING A RANDOM SIGNAL

Following the recent papers by Wax and Kailath (1984) and Bai, Krishnaiah and Zhao (1986), let us consider the model

$$X(t) = A\ s(t) + n(t),\ t = t_1,\ldots,t_n, \qquad (3.1)$$

where $X(t)$ is a p-vector message received at time t, $s(t)$ is a q-vector random signal, $n(t)$ is a p-vector noise component and A is an unknown p×q matrix with elements independent of t. The following assumptions are made regarding the model (3.1).

(i) $s(t_i)$, $i = 1,\ldots,n$, are i.i.d. with the common distribution $N_q(0,\Psi)$, (or $\tilde{N}_q(0,\Psi)$), $q < p$.

(ii) $n(t_i)$, $i = 1,\ldots,n$, are i.i.d. with the common distribution $N_p(0,\sigma^2\Sigma_1)$, (or $\tilde{N}_p(0,\sigma^2\Sigma_1)$).

(iii) $s(t_i)$ and $n(t_j)$ are independent for all i and j.

Under these assumptions

$$X(t) \sim N_p(0,\Sigma_2 = \Gamma+\sigma^2\Sigma_1), \text{ (or } \tilde{N}_p(0,\Sigma_2 = \Gamma+\sigma^2\Sigma_1))$$

where

$$\Gamma = A\Psi A', \text{ (or } A\Psi A^*) \tag{3.2}$$

is of rank $q < p$. Further, if

$$S_2 = \sum_1^n X(t_i)X(t_i)', \text{ (or } \sum_1^n X(t_i)X(t_i)^*) \tag{3.3}$$

then

$$S_2 \sim W_p(n,\Sigma_2), \text{ (or } S_2 \sim \tilde{W}_p(n,\Sigma_2)). \tag{3.4}$$

We consider the problem of testing

$$H_0 : \text{Rank } \Gamma = q<p \quad \text{versus rank } \Gamma \text{ is arbitrary} \tag{3.5}$$

and also of estimating q, the rank of $\Gamma$.

3.1 Case 1 : $\Sigma_1 = I$

In this case $\Sigma_2 = \Gamma+\sigma^2 I$, and it is well-known (see Anderson (1963)) that the likelihood ratio criterion for testing (3.5) is

$$L_q = \frac{(\ell_{q+1}\cdots\ell_p)^s}{[(\ell_{q+1}+\cdots+\ell_p)/(p-q)]^{s(p-q)}} \tag{3.1.1}$$

where $\ell_1 \geq \cdots \geq \ell_p$ are the eigen values of $n^{-1}S_2$ and $s = n/2$ in the real case and $s = n$ in the complex case. In large samples, i.e., as $n \to \infty$

$$-2 \log L_q \sim \chi^2([(p-q)(p-q+1)/2]-1), \text{ (or } \chi^2([(p-q)^2-1]). \tag{3.1.2}$$

As for the estimation of q, Zhao, Krishnaiah and Bai (1986a) suggested a new information theoretic criterion which is more general than those proposed by Akaike (1972), Schwartz (1978) and Rissanen (1978). Their method consists in choosing as an estimate of q the number $\hat{q}$ such that

$$I(\hat{q},C_n) = \min\{I(0,C_n),\ldots,I(p-1,C_n)\} \tag{3.1.3}$$

where

$$I(k,C_n) = -\log L_k + C_n\, \nu(k,p)$$

$$\nu(k,p) = 1 + [k(2p-k+1)/2], \text{(or } 1 + k(2p-k))$$

which is the number of free parameters when rank $\Gamma = k$, and $C_n$ are such that

$$\lim_{n\to\infty} (C_n/n) = 0,$$

$$\lim_{n\to\infty} (C_n/\log\log n) = \infty. \tag{3.1.4}$$

Zhao, Krishnaiah and Bai (1986a) proved that $\hat{q}$ determined as in (3.1.3) is a strongly consistent estimate of q.

3.2 Case 2 : $\Sigma_1 = I$, $\sigma^2=1$

In this case, the likelihood ratio criterion for testing the hypothesis (3.5) is derived by Zhao, Krishnaiah and Bai (1986) as

$$\log L_q = s \sum_{i=1+\min(\tau,q)} (\log \ell_i + 1-\ell_i) \tag{3.2.1}$$

where $s = n/2$ in the real case and $s=n$ in the complex case and $\tau$ is the number of eigen values $\ell_i$ which are greater than unity. The large sample distribution of $-2 \log L_q$ is no longer $\chi^2$. But as suggested by Rao (1983) in a slightly different situation the test of the hypothesis in this case can be carried out in two stages, first as in Case 3.1 taking $\sigma^2$ as unknown, and then examining whether $\sigma^2=1$. However, the criterion of Zhao, Krishnaiah and Bai can be used with (3.2.1) for the estimation of q, i.e., by minimizing

$$-\log L_k + C_n\, \nu(k,p) \tag{3.2.2}$$

where $C_n$ is as in (3.1.4) and

$$\nu(k,p) = [k(2p-k+1)/2], \text{ (or } k(2p-k)). \tag{3.2.3}$$

3.3 Case 3 : $\Sigma_1$ arbitrary and $\sigma^2$ unknown

In addition to $S_2$ defined in (3.3), we suppose that another independent variable $S_1$ is observable and has the distribution

$$S_1 \sim W_p(f,\Sigma_1), \text{ (or } S_1 \sim \tilde{W}_p(f,\Sigma)). \tag{3.3.1}$$

Let $\ell_1 \geq \ldots \geq \ell_p$ be the roots of the equation $|n_2^{-1}S_2 - f^{-1}S_1| = 0$. The likelihood ratio criterion for the hypothesis (3.5) is derived by Rao (1983) in the form

$$-2 \log L_q = -s \log \prod_{i=q+1}^{p} \left[ \left( \frac{n_2 \ell_i + f\hat{\sigma}^2}{n_2+f} \right)^{n_1+n_2} \frac{1}{\ell_i^{n_2} \hat{\sigma}^{2f}} \right] \quad (3.3.2)$$

where s = 1 in the real case and s = 2 in the complex case, and $\hat{\sigma}^2$ is the root of the equation

$$\frac{(p-q)n_2}{f+n_2} = \sum_{i=q+1}^{p} \frac{n_2 m_i}{n_2 m_i + f\sigma^2}. \quad (3.3.3)$$

As $n_2$ and $f \to \infty$, the statistic (3.3.2) is distributed as

$$\chi^2\left(\frac{(p-q)(p-q+1)}{2} - 1\right), \text{ (or } \chi^2((p-q)^2 - 1)). \quad (3.3.4)$$

For estimating q, the method of Zhao, Krishnaiah and Bai is to minimize

$$-\log L_k + C_n \, \nu(k,p) \quad (3.3.5)$$

where $C_n$ is as in (3.1.4) and

$$\nu(k,p) = \frac{k(2p-k+1)}{2} + 1, \text{ (or } k(2p-k) + 1). \quad (3.3.6)$$

3.4 Case 4 : $\Sigma_1$ is arbitrary and $\sigma^2 = 1$

In this case the likelihood ratio criterion is derived by Zhao, Krishnaiah and Bai (1986b) as

$$-2 \log L_q = s \log \prod_{i=1+\min(q,\tau)} \left[ \left( \frac{n_2 \ell_i + f}{n_2+f} \right)^{n_2+f} \ell_i^{-n_2} \right] \quad (3.4.1)$$

where s = 1 for the real case and s = 2 for the complex case, and $\tau$ is the number of $\ell_i$ which are greater than unity. The statistic does not have an asymptotic $\chi^2$ distribution but is useful in the estimation of q. As in other cases we choose q to minimize

$$-\log L_k + C_n \nu(k,p) \tag{3.4.2}$$

where $C_n$ is as in (3.1.4) and

$$\nu(k,p) = \frac{k(2p-k+1)}{2}, \text{ (or } k(2p-k)) \tag{3.4.3}$$

The estimates of q obtained in (3.2.2), (3.3.5) and (3.4.2) are strongly consistent as $n_2$ and $f\to\infty$. Detailed proofs are given in two papers by Zhao, Krishnaiah and Bai (1986a,1986b).

## 4. EXPONENTIAL SIGNAL MODELS

Let $y_t = \xi_t + \eta_t$, t = 1,...,n, be observations taken at equal intervals of time on signals corrupted by noise. The signal $\xi_t$ is considered to be of the form

$$\xi_t = a_1 e^{s_1 t} + \ldots + a_m e^{s_m t} \tag{4.1}$$

where $a' = (a_1,\ldots,a_m)$ and $s' = (s_1,\ldots,s_m)$ are unknown complex vector parameters. Often the value of m, the number of signals (exponential terms in (4.1)), is itself unknown and has to be considered as an essential parameter under estimation. The noise components are taken to be independently distributed with mean zero and common variance $\sigma^2$. The problem has a long history starting with the pioneering work of Prony (1795) two hundred years ago. In a series of papers Tufts and Kumaresan (see Kumaresan

(1982), Tufts and Kumaresan (1982) and Kumaresan and Tufts (1982) and the numerous references therein) suggested some new approaches to the problem based on Prony's parametrization of the signal process described below.

Prony (1795) observed that $\xi_i$ as defined in (4.1) satisfy the recurrence relations

$$\xi_{i+m+1} + g_1\xi_{i+m} + \dots + g_m\xi_i = 0, \ i=1,\dots,n-m-1, \qquad (4.2)$$

where $g' = (g_m,\dots,g_1,1)$ is a function of s, which may be regarded as an alternative parameter to s. The equations (4.2) lead to the observational equations

$$-y_{i+m+1} = g_1 y_{i+m} + \dots g_m y_i, \ i=1,\dots,n-m-1 \qquad (4.3)$$

which can be written in the matrix form

$$GY = 0, \text{ with } Y' = (y_1,\dots,y_n) \qquad (4.4)$$

choosing the (n-m-1) x n matrix G appropriately with each row containing the row vector g' and a number of zeros, the position of g' being shifted by one element when we go from one row to the next row. Much of the previous work is centered on the equations (4.3, 4.4) and the estimation of g by minimizing

$$Y^*G^*GY \text{ or } Y^*G^*GY/g'g \qquad (4.5)$$

fixing an appropriate value for m. Once g is estimated, $\exp(s_i)$ are obtained as the roots of the polynomial equation

$$0 = 1 + g_1 z^{-1} + \dots + g_m z^{-m} = \prod_{n=1}^{m} (1-z^{-r} e^{s_r}) \qquad (4.6)$$

and a is estimated by minimizing

$$(Y-xa)^*(Y-xa) \tag{4.7}$$

taking

$$x \stackrel{\Delta}{=} \begin{pmatrix} \exp(\hat{s}_1) & \dots & \exp(\hat{s}_m) \\ . & \dots & . \\ \exp(\hat{s}_1 n) & \dots & \exp(\hat{s}_m n) \end{pmatrix} \tag{4.8}$$

as fixed. Tufts and Kumaresan in the papers cited above make some refinements by starting with a larger value, say $\ell > m$, and use the extra terms to reduce the noise part of the model. However, it is not clear whether minimizing (4.5) ignoring the correlations between the components of GY lead to consistent estimators of the unknown parameters. It may be *noted that the covariance matrix* of GY as defined in (4.4) is $\sigma^2 GG^*$, in which case the appropriate quadratic form to be minimized is

$$Y^* G^* (GG^*)^{-1} GY \tag{4.9}$$

and not (4.4), although the minimization problem associated with (4.9) is far more complicated.

It is shown (see Smyth (1985) and references therein) that the estimates of a and s obtained from (4.6) and (4.7) using the g estimated from (4.9) are, indeed, maximum likelihood estimates, i.e., those obtained by minimizing

$$\sum_{t=1}^{n} |y_t - a_1 e^{s_1 t} - \dots - a_m e^{s_m t}|^2 \tag{4.10}$$

with respect to $a_i$ and $s_j$, under the assumption that the error terms are normally distributed. We can also look upon the estimates of $a_i$ and $s_i$

obtained by minimizing (4.10) as non-linear least squares estimates without any distributional assumptions. Smyth (1985) has developed an efficient algorithm for obtaining g which minimizes (4.9), and then estimating s and a through the steps in (4.6) and (4.7), which solves the non-linear least squares problem (4.10) for given m.

The choice of m can be made by a suitable information theoretic model selection criterion such as the one used in the previous section. If the minimum value of (4.10) is denoted by $R_m$, then under the assumption of normality of $\eta_t$, the error components, the information theoretic criterion takes the form

$$n \log R_m + C_n(4m) \tag{4.11}$$

where $C_n$ are chosen to satisfy the conditions (3.1.4). The AIC criterion (Akaike (1972)) corresponds to the choice $C_n = 1$, which may be sufficient to provide a fairly accurate estimate of m.

However, a more relevant and satisfactory method in finite samples is provided by the cross validation (CV) approach, although the computations may be extremely heavy. (See for instance papers by Rao (1984) and Rao and Boudreau (1985) for such an approach in a prediction problem.)

In the CV method we leave one of the values, say $y_i$, but replace it by a variable $Y_i$. For any choices of $Y_i$ and m, using Smyth's algorithm, we compute

$$R(Y_i,m) = \min_{a,s} \sum_{t=1}^{n} |y_t - a_1 e^{s_1 t} - \dots - a_m e^{s_m t}|^2 \tag{4.12}$$

where $a' = (a_1,\dots,a_m)$ and $s' = (s_1,\dots,s_m)$. Then for given m, by a suitable computer program, we find $\hat{y}_{im}$ such that

$$R(\hat{y}_{im},m) = \min_{Y_i} R(Y_i,m) \tag{4.13}$$

which provides $\hat{y}_{im}$ as an estimate of $Y_i$ for given m. Then comparing $\hat{y}_{im}$ with the observed $y_i$, the cross validation error (CVE) is obtained as

$$R_*(m) = \sum_{i=1}^{n} (y_i - \hat{y}_{im})^2, \tag{4.14}$$

and finally m is chosen as that value for which $R_*(m)$ is a minimum.

As observed earlier, the computations involved in the above procedure are extremely heavy. However, simplication may be effected in some ways.

1. If n is large, we may choose every alternative or every third value among the components of $(y_1,\ldots,y_n)$ for cross validation. This cuts down on the number of terms in (4.14) and reduces the computing time considerably.

2. It may be noted that $\hat{y}_{im}$ can be computed by an alternative method as

$$\hat{y}_{im} = \sum_{j=1}^{m} \hat{a}_j e^{\hat{s}_j i} \tag{4.15}$$

where $\hat{a}_j$, $\hat{s}_j$, $j=1,\ldots,m$, are the values minimizing the expression,

$$\sum_{t=1}^{i-1} \left| y_t - \sum_{1}^{m} a_j e^{s_j t} \right| + \sum_{t=i+1}^{n} \left| y_t - \sum_{1}^{m} a_j e^{s_j t} \right|^2. \tag{4.16}$$

Due to the absence of the term $y_i$ in (4.16), one cannot take full advantage of Prony's reparametrization. But if the optimum values $\hat{a}_j$ and $\hat{s}_j$ can be found directly through some other algorithm, then $\hat{y}_{im}$ can be obtained as in (4.15).

3. We defined $R(Y_i, m)$ as the minimum of the expression on the right hand side of (4.12). For purposes of estimating m, one could use an approximation

$$R(Y_i, m) = \sum_{t=1}^{m} |y_t - \hat{a}_1 e^{\hat{s}_1 t} - \ldots - a_m e^{\hat{s}_m t}|^2 \qquad (4.17)$$

where $a_i$ and $s_j$ are estimates obtained by methods such as those suggested by Tufts and Kumaresan.

## REFERENCES

1. Akaike, H. (1972). Information theory and an extension of the maximum likelihood principle. In Proc. Second International Symposium on Information Theory. Supp. to Problems of Information Theory and Control, 267-281.

2. Anderson, T.W. (1963). Asymptotic theory for principal component analysis, Ann. Math. Statist.. 34, 122-148.

3. Bai, Z.D., Krishnaiah, P.R. and Zhao, L.C. (1986). Signal processing using model selection methods. Tech. Rept. 86-03, Center for Multivariate Analysis, Pittsburgh.

4. Khatri, C.G. and Rao, C. Radhakrishna (1985a). Effects of estimated noise covariance matrix in optimal signal detection. Tech. Rept. 85-38, Center for Multivariate Analysis, Pittsburgh.

5. Khatri, C.G. and Rao, C. Radhakrishna (1985b). Test for a specified signal when the noise covariance matrix is unknown. Tech. Rept. 85-47, Center for Multivariate Analysis, Pittsburgh.

6. Khatri, C.G., Rao, C.R. and Sun, Y.N. (1986). Tables for obtaining confidence bounds for realized signal to noise ratio with an estimated discriminant function. Communications in Statistics, Simulation and Computation 15, 1-14.

7. Kumaresan, Ramdas (1982). Estimating the parameters of exponentially damped or undamped sinusoidal signals in noise. Ph.D. thesis, University of Rhode Island.

8. Kumaresan, R. and Tufts, D.W. (1982). Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. IEEE Trans. Acoust., Speech, Signal Processing ASSP-30, 833-840.

9. Prony, R. de (1795). Essai experimentale et analytique. J. Ecole Polytechnique (Paris), 24-76.

10. Rao, C. Radhakrishna (1946). Tests with discriminant functions in multivariate analysis. Sankhya 7, 407-414.

11. Rao, C. Radhakrishna (1948). The utilization of multiple measurements in problems of biological classification. J. Roy. Statist. Soc. B10, 159-203.

12. Rao, C. Radhakrishna (1949). On some problems arising out of discrimination with multiple characters. Sankhya 9, 343-364.

13. Rao, C. Radhakrishna (1973). Linear Statistical Inference and its Applications (Second Edition), Wiley, New York.

14. Rao, C. Radhakrishna (1983). Likelihood ratio tests for relationships between two covariance matrices. In Studies in Econometrics, Time Series, and Multivariate Statistics (Eds. S. Karlin, T. Amemiya and L.A. Goodman), Academic Press, 529-544.

15. Rao, C. Radhakrishna (1984). Prediction of future observations in polynomial growth curve models. Proc. Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions, 512-520.

16. Rao, C. Radhakrishna and Boudreau, R. (1985). Prediction of future observations in a factor analytic type growth model. In Multivariate Analysis VI (Ed. P.R. Krishnaiah), Elsevier Science Publishers, 449-466.

17. Reed, I.S., Mallet, J.D. and Brennan, L.E. (1974). Rapid convergence rates in adaptive arrays. IEEE Transactions on Aerospace and Electronic Systems. 10, 853-863.

18. Rissanen, J. (1978). Modeling by shortest data description. Automatica 14, 465-471.

19. Schwartz, G. (1978). Estimating the dimension of a model. Ann. Statist. 6, 461-464.

20. Smyth, G.K. (1985). Coupled and separable iterations in non-linear estimation. Ph.D. thesis, Australian National University.

21. Tufts, D.W. and Kumaresan, R. (1982). Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. Proc. IEEE, Special Issue on Spectral Estimation 70, 975-989.

22. Wax. M. and Kailath, T. (1985). Detection of signals by information theoretic criteria. IEEE Trans. Accoustics, Speech, and Signal Processing ASSP-33, 387-392.

23. Zhao, L.C., Krishnaiah, P.R. and Bai, Z.D. (1986a). On the detection of the number of signals in the presence of white noise. J. Multivariate Analysis 20.

24. Zhao, L.C., Krishnaiah, P.R. and Bai, Z.D. (1986b). On the detection of number of signals when the noise covariance matrix is arbitrary. J. Multivariate Analysis 20.

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER 86-23 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) Some recent results in signal detection | | 5. TYPE OF REPORT & PERIOD COVERED Technical - September 1986 |
| | | 6. PERFORMING ORG. REPORT NUMBER 86-23 |
| 7. AUTHOR(s) C. Radhakrishna Rao | | 8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0008 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis University of Pittsburgh 512 Thackeray Hall, Pittsburgh, PA 15260 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Dept. of the Air Force Bolling Air Force Base, DC 20332 | | 12. REPORT DATE September 1986 |
| | | 13. NUMBER OF PAGES 22 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report) unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Discriminant function, exponential signal models, information criteria in model selection, Prony's method, signal processing.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Some recent results on the detection and estimation of signals in the presence of noise are discussed. An exact confidence lower bound is obtained for the discriminatory power of an estimated linear discriminant function for signal detection. Information theoretic criteria are suggested for the estimation of signals. A new method is proposed for determining the number of signals and estimating them in exponential signal models.

DD FORM 1 JAN 73 1473

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

4-[illegible]-87

DTIC